基于深度强化学习的光储充电站储能系统优化运行

陈亭轩,徐潇源,严 正,朱彦名
(上海交通大学 电力传输与功率变换控制教育部重点实验室,上海 200240)

摘要:优化储能充放电策略有利于提升光储充电站运行经济性,但是现有模型驱动的随机优化方法无法全面考虑储能系统的复杂运行特性以及光伏发电功率、电动汽车充电负荷的不确定性。因此,提出一种基于深度强化学习的光储充电站储能系统全寿命周期优化运行方法。首先对储能运行效率模型和容量衰减模型进行精细化建模。然后考虑电动汽车充电需求、光伏出力和电价的不确定性,在满足电动汽车充电需求和光伏消纳的条件下,以光储充电站收益最大化为目标,建立了基于强化学习的储能优化运行问题。考虑到储能充放电决策动作的连续性,采用双延迟深度确定性策略梯度算法进行求解。采用实际历史数据对模型进行训练,根据当前时段状态对储能充放电策略进行实时优化。最后,对所提方法及模型进行测试,并将所提出的方法与传统模型驱动方法进行对比,结果验证了所提方法及模型的有效性。

关键词:储能;光储充电站;不确定性;深度强化学习;优化
中图分类号:TM 715;U 469.72 **文献标志码**:A DOI:10.16081/j.epae.202110037

0 引言

电动汽车EV(Electric Vehicle)的普及是实现碳中和的重要途经,也是应对全球能源问题和环境危机的一种有效解决方案[1]。但目前我国发电侧能源仍然以煤炭为主,如果只是通过普通充电站对电动汽车进行充电,所消耗的电能大部分仍来自煤炭等非清洁能源发电,在节能减排上并不具备明显优势[2]。将新能源作为电动汽车充电站的主要电源,能够实现真正意义上的"低碳"。

光伏充电站是利用光伏为电动汽车充电的新能源充电站。由于光伏出力与天气因素密切相关,电动汽车充电需求也具有较大的不确定性,光伏充电站通常配置储能设备,实现能量的储存和调节,提高光伏的消纳率,同时通过"低储高发"获取经济利润[3]。目前,在光伏充电站储能系统的调度策略方面已有一些研究成果。文献[4]通过优化储能有功和无功功率应对电动汽车充电负荷的不确定性。文献[5]针对同时配备光伏、燃料电池和储能的电动汽车充电站,以降低充电站的运行成本及其对配电网的影响为目标,采用随机动态规划方法求解充电站调度问题。文献[6]以最小化电动汽车充电站的运行成本和储能循环电量为目标,提出了电动汽车充电时间和电池储能充放电功率的优化方法。文献[3-5]采用随机优化及鲁棒优化方法应对光伏充电站运行中电动汽车充电负荷、光伏出力等不确定性因素的影响。但是随机优化依赖准确的概率分布模型,优化结果的样本外表现较差且大规模随机优化问题规模较大;鲁棒优化因考虑最劣场景下的调度方案,所得结果可能较为保守[7]。文献[8]采用分布鲁棒优化方法处理光伏充电站选址定容中的不确定性因素,该方法虽然解决了随机优化中模型不准确和鲁棒优化中结果保守的问题,但是计算量较大,应用于实时调度时受到一定限制。

另一方面,储能的运行效率和使用寿命直接影响光伏充电站经济收益,因此考虑精细化的储能运行模型具有重要意义。现有文献多采用简化的运行模型进行储能优化问题的求解。部分文献将储能的运行效率视为恒定值[4,9]。部分文献仅着重考虑某个因素对储能寿命衰减的影响,文献[6,10]分别从循环电量和放电深度DOD(Depth Of Discharge)两方面考虑储能寿命的衰减;文献[11]认为储能的使用寿命与吞吐量密切相关,并通过限制日吞吐量延长使用时间。实际上,电化学储能的运行效率和寿命衰减受到运行温度、充放电功率、DOD和荷电状态SOC(State Of Charge)等因素影响[12]。文献[13]将电池内电阻表示为电池SOC的非线性函数,通过计算电阻损耗建立储能运行效率模型;文献[14]对由多个电池单体组成的储能设备的运行效率进行二项式拟合,得到充放电功率和SOC的函数;文献[15]考虑DOD和循环次数对储能使用寿命的影响,通过雨流计数法和等效循环寿命法对电池寿命进行估计;文献[16]考虑多种因素对储能寿命衰减的影响,提出了一种半经验的锂离子电池退化模型,采用雨流计数法计算电池寿命损失。然而,在现有的光储充电站优化运行问题中,文献[14,16]中精细化的、非线性的储能模型会增加优化问题求解的复杂度。因此,光伏发电、电动汽车充电负荷等不确定性因素

收稿日期:2021-06-05;**修回日期**:2021-09-07
基金项目:国家自然科学基金资助项目(52077136)
Project supported by the National Natural Science Foundation of China(52077136)

的复杂特征和储能系统的非线性运行特点显著增加了基于解析数学模型的优化问题求解难度。

深度强化学习DRL(Deep Reinforcement Learning)是一种数据驱动的决策问题求解方法，具有如下特点：①具有自适应性、无模型性和从历史数据中学习的能力；②能够在复杂环境下学习到较好的控制策略[17]。因此，针对考虑不确定性因素和复杂非线性模型的储能系统优化运行问题，DRL的上述特点使得其在学习储能充放电策略中具有很大的应用潜力。文献[17]考虑电价不确定性以及储能模型的非线性，采用DRL解决储能套利问题；文献[18]采用DRL优化电动汽车充电策略，但是其未考虑储能及其详细模型。综上可知，采用DRL解决不确定性环境下光储充电站优化运行问题的研究相对较少。

本文以光储充电站的收益最大化为目标，详细考虑光储充电站中储能运行效率和寿命衰减过程，以及光伏发电出力、电动汽车充电负荷、电价的不确定性，构建储能系统优化运行的DRL问题，并采用双延迟深度确定性策略梯度算法进行求解。最后，采用实际数据进行算例分析，并与基于解析数学模型的优化方法求解结果进行对比，以验证所提模型和方法的有效性。

1　光储充电站运行问题

本文以光储充电站为研究对象，光储充电站由光伏发电、储能和电动汽车充电桩组成。光储充电站运行方式如图1所示。光伏发电装置向系统提供能量，满足电动汽车的充电需求，储能可储存多余电能，也可在光伏不足时为电动汽车提供能量。光储充电站可向电网卖出和购入电能，以满足电动汽车充电需求，促进光伏并网消纳。

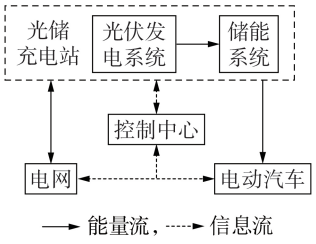


图1　光储充电站运行方式

Fig.1　Operation mode of photovoltaic-storage charging station

1.1　目标函数

在消纳光伏和满足电动汽车充电需求的前提下，优化问题目标为最大化光储充电站的经济收益。光储充电站可通过为电动汽车充电、与电网进行交易获取经济收益，同时还需考虑储能系统的使用寿命，因此目标函数包括充电收益、电网交易收益和储能容量衰减成本，如式(1)所示。

$$f=\max\sum_{t=1}^{T_0}(B_{\mathrm{ev},t}+B_{\mathrm{grid},t}-C_{\mathrm{batt},t}) \tag{1}$$

$$B_{\mathrm{ev},t}=\lambda_{\mathrm{ev}}P_{\mathrm{ev},t}\Delta t \tag{2}$$

$$B_{\mathrm{grid},t}=\begin{cases}\lambda_{\mathrm{s},t}P_{\mathrm{grid},t}\Delta t & P_{\mathrm{grid},t}\geqslant 0\\ \lambda_{\mathrm{b},t}P_{\mathrm{grid},t}\Delta t & P_{\mathrm{grid},t}<0\end{cases} \tag{3}$$

式中：$B_{\mathrm{ev},t}$为t时段光储充电站为电动汽车提供电能的充电收益；$B_{\mathrm{grid},t}$为t时段光储充电站与电网交易的收益；$C_{\mathrm{batt},t}$为t时段光储充电站中储能容量衰减成本；T_0为一个周期内的时段数，本文考虑一个周期为1 d，共有96个时段；$P_{\mathrm{ev},t}$为t时段光储充电站的电动汽车充电需求总功率；λ_{ev}为电动汽车充电价格；Δt为单位时段时长；$P_{\mathrm{grid},t}$为t时段光储充电站与电网交易的功率，其值为正、负分别表示光储充电站向电网售电、光储充电站向电网购电，为0表示既不售电也不购电；$\lambda_{\mathrm{s},t}$和$\lambda_{\mathrm{b},t}$分别为t时段光储充电站向电网售、购电的价格。

1.2　约束条件

1)功率平衡约束。

$$P_{\mathrm{solar},t}+P_{\mathrm{es},t}=P_{\mathrm{ev},t}+P_{\mathrm{grid},t} \tag{4}$$

式中：$P_{\mathrm{solar},t}$为光伏在t时段的出力；$P_{\mathrm{es},t}$为储能在t时段的出力，本文假设$P_{\mathrm{es},t}$在t时段内保持不变。

2)储能约束。

(1)储能出力上下限约束。

$$-P_{\mathrm{es,max}}\leqslant P_{\mathrm{es},t}\leqslant P_{\mathrm{es,max}} \tag{5}$$

式中：$P_{\mathrm{es,max}}$为储能出力的最大值。

(2)储能SOC约束。

$$S_{\mathrm{SOC}}^{\min,t}\leqslant S_{\mathrm{SOC}}^{t}\leqslant S_{\mathrm{SOC}}^{\max,t} \tag{6}$$

式中：S_{SOC}^{t}为t时段储能的SOC；$S_{\mathrm{SOC}}^{\max,t}$、$S_{\mathrm{SOC}}^{\min,t}$分别为$t$时段储能的最大、最小SOC。

储能电池的SOC转换可由式(7)表示。

$$S_{\mathrm{SOC}}^{t+1}=\begin{cases}S_{\mathrm{SOC}}^{t}-\eta_t^{\mathrm{ch}}\dfrac{P_{\mathrm{es},t}\Delta t}{E_{\mathrm{ini}}} & P_{\mathrm{es},t}\leqslant 0\\ S_{\mathrm{SOC}}^{t}-\dfrac{1}{\eta_t^{\mathrm{dis}}}\dfrac{P_{\mathrm{es},t}\Delta t}{E_{\mathrm{ini}}} & P_{\mathrm{es},t}>0\end{cases} \tag{7}$$

式中：E_{ini}为储能电池衰减前的最大容量；η_t^{ch}和η_t^{dis}分别为t时段储能电池的充电和放电效率。

(3)储能末时段的SOC约束。

为保证储能能够在下一个控制周期正常工作，应使得末时段储能动作之后的SOC为给定值，如式(8)所示。

$$S_{\mathrm{SOC}}^{T_0+1}=S_{\mathrm{SOC}}^{1} \tag{8}$$

2　储能模型

2.1　储能运行效率模型

电池的充放电效率可采用电池稳态电路等效模型计算，单体电池的稳态电路如图2所示。图中，V_{oc}为开路电压；I_{batt}为电池电流；P_{out}为输出功率；R_1为串联电阻；R_2和R_3分别为与暂态响应相关的短时间响

应电阻和长时间响应电阻。上述变量都与电池SOC S_{SOC}具有非线性关系[19]，如式(9)所示。

$$\begin{cases} V_{oc}=g_0e^{-g_1S_{SOC}}+g_2+g_3S_{SOC}-g_4S_{SOC}^2+g_5S_{SOC}^3 \\ R_1=b_0e^{-b_1S_{SOC}}+b_2+b_3S_{SOC}-b_4S_{SOC}^2+b_5S_{SOC}^3 \\ R_2=c_0e^{-c_1S_{SOC}}+c_2 \\ R_3=d_0e^{-d_1S_{SOC}}+d_2 \\ R_{oc}=R_1+R_2+R_3 \end{cases} \tag{9}$$

式中：R_{oc}为终端电阻；g_0—g_5、b_0—b_5、c_0—c_2和d_0—d_2均为系数。

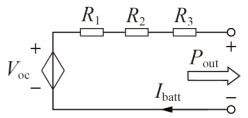


图2 稳态电路等效模型

Fig.2 Equivalent model of static circuit

对于特定的SOC和P_{out}，可以通过求解式(10)得到流过单体电池的电流。

$$I_{batt}^2R_{oc}-I_{batt}V_{oc}+P_{out}=0 \tag{10}$$

单体电池的充电和放电效率可由式(11)表示。

$$\begin{cases} \eta^{ch}=\dfrac{V_{oc}}{V_{oc}-I_{batt}R_{oc}} \\ \eta^{dis}=\dfrac{V_{oc}-I_{batt}R_{oc}}{V_{oc}} \end{cases} \tag{11}$$

本文所采用的储能电池由多个单体电池串并联组成，其等效电流和充放电效率的计算方式分别为：

$$I_{eq}^2R_{oc}-IV_{oc}+P_{es}\frac{1}{N_{series}N_{para}}=0 \tag{12}$$

$$\begin{cases} \eta^{ch}=\dfrac{V_{oc}}{V_{oc}-I_{eq}R_{oc}} \\ \eta^{dis}=\dfrac{V_{oc}-I_{eq}R_{oc}}{V_{oc}} \end{cases} \tag{13}$$

式中：I_{eq}为储能电池电流；I为多个单体电池串并联后的等效电流；N_{series}为串联的电池单元数量；N_{para}为并联的电池单元数量。

根据式(13)计算运行效率，并通过二阶多项式拟合运行效率与输出功率、SOC的关系[14]：

$$\begin{cases} \eta^{ch}=e_0+e_1S_{SOC}+e_2S_{SOC}^2+e_3P_{es}+e_4P_{es}^2+e_5P_{es}S_{SOC} \\ \dfrac{1}{\eta^{dis}}=h_0+h_1S_{SOC}+h_2S_{SOC}^2+h_3P_{es}+h_4P_{es}^2+h_5P_{es}S_{SOC} \end{cases} \tag{14}$$

式中：e_0—e_5和h_0—h_5为系数。

2.2 储能衰减模型

电池容量的衰减是环境温度、DOD、SOC以及电池运行时间的函数。电池容量衰减并非是每一个循环周期衰减容量的简单求和，而是一个非线性过程。因此本文采用一种半经验电池容量衰减模型[16]。N个循环周期内的电池容量退化模型如式(15)所示。

$$E_{loss}=\begin{cases} [1-\alpha_{sei}e^{-N\beta_{sei}f_d}-(1-\alpha_{sei})e^{-Nf_d}]E_{ini} & E'_{loss}=0 \\ [1-(1-E'_{loss})e^{-Nf_d}]E_{ini} & E'_{loss}>0 \end{cases} \tag{15}$$

式中：E'_{loss}、E_{loss}分别为N个循环周期前、后的衰减容量，$E'_{loss}=0$表示该电池尚未进行任何充放电动作；α_{sei}和β_{sei}为新电池使用时固体电解质界面SEI(Solid Electrolyte Interface)膜形成时的影响系数；f_d为单个循环周期电池容量衰减函数，其可表示为温度T、DOD δ、SOC以及时间t_1的函数，如式(16)所示。

$$f_d=(f_\delta(\delta)+f_t(t_1))f_{SOC}(S_{SOC})f_T(T) \tag{16}$$

$$\begin{cases} f_T(T)=e^{k_T\frac{(T-T_{ref})T_{ref}}{T}} \\ f_{SOC}(S_{SOC})=e^{k_{SOC}(S_{SOC}-S_{SOCref})} \\ f_t(t_1)=k_tt_1 \\ f_\delta(\delta)=(k_{\delta1}\delta^{k_{\delta2}}+k_{\delta3})^{-1} \end{cases} \tag{17}$$

式中：$f_\delta(\delta)$、$f_t(t_1)$、$f_{SOC}(S_{SOC})$和$f_T(T)$分别为DOD、时间、SOC和温度对储能电池寿命的影响函数；T_{ref}为参考温度；S_{SOCref}为参考SOC；k_T、k_{SOC}、k_t、$k_{\delta1}$—$k_{\delta3}$为影响因素的相关系数。

储能电池的剩余可用容量E_{re}为：

$$E_{re}=E_{ini}-E_{loss} \tag{18}$$

式(15)—(17)对电池容量衰减的计算方式需要每个循环周期的参数，然而在实际运行中，电池通常采取不规则的充放电行为，导致循环周期及其参数难以直接获得[20]。本文采用疲劳分析中的雨流计数法计算电池的循环周期及其相应参数。基于雨流计数法的储能电池容量衰减计算流程如图3所示[16]。

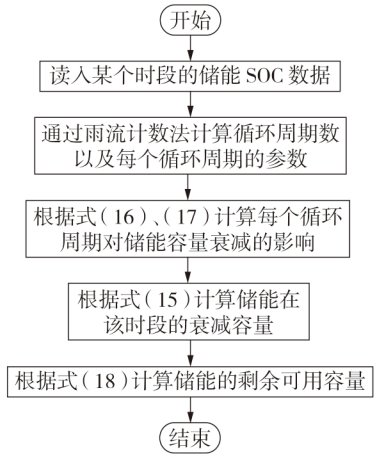


图3 储能电池容量衰减计算流程

Fig.3 Flowchart of calculating battery capacity degradation

3 基于DRL的储能优化运行方法

3.1 强化学习问题建模

采用强化学习解决光储充电站优化运行问题时，需要根据原有优化问题设计对应的状态空间、动作空间和奖惩函数。

1）状态空间。

对于任意t时段，状态空间由光伏发电功率、电动汽车总充电需求功率、电价和储能SOC共同构成，如式（19）所示。

$$s_t=\left\{P_{\text{solar},t},P_{\text{ev},t},\lambda_{\text{grid},t},S_{\text{SOC}}^{t}\right\} \tag{19}$$

式中：$\lambda_{\text{grid},t}$为t时段的实时电价，本文假设光储充电站与电网的交易电价均为实时电价。

2）动作空间。

将储能出力设置为动作空间，通过约束条件式（5）限制动作空间范围，如式（20）所示。

$$a_t=\begin{cases}-P_{\text{es,max}} & P_{\text{es},t}<-P_{\text{es,max}}\\ P_{\text{es,max}} & P_{\text{es},t}>P_{\text{es,max}}\\ P_{\text{es},t} & -P_{\text{es,max}}\leqslant P_{\text{es},t}\leqslant P_{\text{es,max}}\end{cases} \tag{20}$$

式中：a_t为t时段的储能出力动作。

为保证系统始终满足功率平衡约束，光储充电站系统与电网交互的功率由式（4）计算得到，不作为动作空间的一部分，即：

$$P_{\text{grid},t}=P_{\text{solar},t}+P_{\text{es},t}-P_{\text{ev},t} \tag{21}$$

3）奖惩函数。

奖惩函数决定环境对某一时段储能充放电动作的即时回报，其影响强化学习智能体对动作的选择。本文中的奖惩函数由收益奖励、储能容量衰减惩罚和约束条件惩罚组成。

（1）收益奖励。

收益奖励与目标函数式（1）中光储充电站获取的充电收益以及电网交易收益对应，如式（22）所示。

$$r_{t,1}=B_{\text{ev},t}+B_{\text{grid},t} \tag{22}$$

式中：$r_{t,1}$为t时段的收益奖励。

（2）储能容量衰减惩罚。

储能容量衰减惩罚与目标函数式（1）中的储能衰减成本$C_{\text{batt},t}$对应。采用图3所示方法能够计算得到一段时间内的储能容量衰减成本，但其不能作为强化学习中的即时回报；因此，采用式（23）计算一段时间内单位功率对应的储能容量衰减成本，得到相应惩罚因子α_k，且经过一个容量衰减计数周期T_1后更新α_k[17]。

$$\alpha_k=\frac{E_{\text{re},k}^{\text{start}}-E_{\text{re},k}^{\text{end}}}{\sum_{t=1}^{T_1}\left|P_{\text{es},t}\right|}c_{\text{batt}} \tag{23}$$

式中：$E_{\text{re},k}^{\text{start}}$和$E_{\text{re},k}^{\text{end}}$分别为储能在第$k$个容量衰减计数周期的剩余容量初始值和结束值，均通过图3所示的储能电池容量衰减计算流程得到；c_{batt}为成本系数。

因此，储能容量衰减的即时惩罚为：

$$r_{t,2}=\alpha_k\left|P_{\text{es},t}\right| \tag{24}$$

式中：$r_{t,2}$为t时段储能容量衰减惩罚。

（3）SOC上下限惩罚。

由于储能动作对下一时段的最大可用容量有影响，不能直接限定动作范围，因此对使得S_{SOC}^{t+1}超出上、下限的动作进行惩罚：

$$r_{t,3}=\begin{cases}\left|a_t-\dfrac{S_{\text{SOC}}^{t}-S_{\text{SOC}}^{\min,t+1}}{\eta_t^{\text{ch}}\Delta t}\right| & S_{\text{SOC}}^{t+1}<S_{\text{SOC}}^{\min,t+1}\\ 0 & S_{\text{SOC}}^{\min,t+1}\leqslant S_{\text{SOC}}^{t+1}\leqslant S_{\text{SOC}}^{\max,t+1}\\ \left|a_t-\eta_t^{\text{dis}}\dfrac{S_{\text{SOC}}^{t}-S_{\text{SOC}}^{\max,t+1}}{\Delta t}\right| & S_{\text{SOC}}^{t+1}>S_{\text{SOC}}^{\max,t+1}\end{cases} \tag{25}$$

式中：$r_{t,3}$为对超出SOC上、下限的惩罚。

（4）储能末时段SOC惩罚。

根据约束条件式（8）得到储能末时段SOC惩罚为：

$$r_{t,4}=\begin{cases}\left|S_{\text{SOC}}^{T_0+1}-S_{\text{SOC}}^{1}\right| & t=T_0\\ 0 & t<T_0\end{cases} \tag{26}$$

式中：$r_{t,4}$为储能末时段的SOC惩罚。

综上所述，强化学习的奖惩函数为：

$$r_t=\sigma_1 r_{t,1}-\sigma_2 r_{t,2}-\sigma_3 r_{t,3}-\sigma_4 r_{t,4} \tag{27}$$

式中：r_t为t时段的即时回报；σ_1—σ_4为各部分奖惩的权重系数，且均为正数。

3.2　双延迟深度确定性策略梯度算法

强化学习智能体（Agent）根据当前状态s_t，按照策略从动作空间中选择动作a_t，并根据奖惩函数获取即时奖励$r_t(s_t,a_t)$。本文考虑一个学习任务共有T_0个时段，从t时段到学习任务结束的累积奖赏R_t为：

$$R_t=\sum_{i=t}^{T_0}\gamma^{i-t}r_i(s_i,a_i) \tag{28}$$

式中：γ为折扣因子，决定未来奖赏对累积奖赏的影响。

强化学习的目标函数J是寻找最优策略π使得智能体在T_1内的期望累积奖赏最大，即：

$$\max J=E_{s_i\sim\rho^{\pi},a_i\sim\pi}[R_1] \tag{29}$$

式中：ρ为某一策略下的状态转移概率分布；R_1为整个学习任务的累积奖赏，可由式（28）得到。

状态-动作值函数$Q(s_t,a_t)$表示在策略π下产生的长期回报期望，如式（30）所示。

$$Q(s_t,a_t)=E_{a_t\sim\pi}\left[R_t\middle|s_t,a_t\right] \tag{30}$$

状态-动作值函数的贝尔曼方程表示为：

$$Q(s_t,a_t)=r_t+\gamma E_{a_{t+1}\sim\pi}[Q(s_{t+1},a_{t+1})] \tag{31}$$

Q学习（Q-learning）算法是一种基于值函数的强化学习算法，采用式（31）进行迭代更新，当Q值收敛于最优值Q^*后，可得到状态为s的最优动作a^*，如式（32）所示。

$$a^*=\underset{a\in A}{\operatorname{argmax}}\,Q^*(s,a) \tag{32}$$

式中：A为动作空间。

式（32）要求动作空间离散化，这会导致动作空间维数过大和次优解的问题，因此Q学习算法难以

应用于储能充放电的连续空间的决策问题。演员-评论家AC(Actor-Critic)框架采用神经网络克服Q学习算法所面临的困难:采用actor网络拟合状态和动作的映射关系,避免动作空间离散化;采用critic网络对Q值函数进行拟合,评估actor网络的策略,使输出动作逼近最优解。因此本文基于AC框架的强化学习算法适用于连续动作空间的决策问题。

在AC框架中,actor网络根据某一策略将当前状态映射到某指定动作,如式(33)所示。

$$a_t=\mu\left(s_t\middle|\theta_\mu\right)+N_t \tag{33}$$

式中:$\mu\left(s\middle|\theta_\mu\right)$为状态-动作映射关系的拟合函数,$\theta_\mu$为actor网络参数;$N_t$为噪声。

actor网络通过策略梯度更新网络参数[21],如式(34)所示。

$$\nabla_{\theta_\mu}J\approx E_{s_t\sim\rho^\beta,a_t\sim\beta}\left[\nabla_{\theta_\mu}Q\left(s,a\middle|\theta_q\right)\middle|_{s=s_t,a=\mu\left(s_t\middle|\theta_\mu\right)}\right] \tag{34}$$

式中:β为某一随机策略;$Q\left(s,a\middle|\theta_q\right)$为critic网络评估的拟合函数;$\theta_q$为critic网络参数。

critic网络通过式(31)对在状态s_t下选择的动作a_t进行评价,通过最小化损失函数更新网络参数。损失函数L_Q为:

$$L_Q=E_{s_t\sim\rho^\beta,a_t\sim\beta,r_t\sim E}\left[\left(y_t-Q\left(s_t,a_t\middle|\theta_q\right)\right)^2\right] \tag{35}$$

式中:y_t为Q值估计。

状态-动作值函数$Q(s,a)$的估计值y_t是建立在对后续状态的估计值之上的,因此存在误差。误差的累积可能会导致较大的Q值高估偏差[19]。由于actor网络更新也与价值函数有关,过高地估计Q值还可能使得网络收敛于次优策略。双延迟深度确定性策略梯度TD3(Twin Delayed Deep Deterministic Policy Gradient)算法在AC框架的基础上改善了critic网络$Q\left(s,a\middle|\theta_q\right)$过高估计的问题[22]。

1)TD3算法采用2个critic网络$Q_1\left(s,a\middle|\theta_q\right)$和$Q_2\left(s,a\middle|\theta_q\right)$对actor的动作值函数进行估计,并使用二者之中的较小值作为估计值,如式(36)所示。

$$y_t=r+\gamma\min_{i=1,2}\left(Q_i\left(s_{t+1},\mu(s_{t+1})\middle|\theta_{q_i}+\varepsilon\right)\right) \tag{36}$$

式中:$Q_i\left(s,a\middle|\theta_{qi}\right)$为第$i$个critic网络的拟合函数;$\varepsilon$为噪声,作用是平滑$Q$值估计。

2)为提高算法的稳定性和收敛性,TD3也采用actor目标网络(target network)和critic目标网络,其结构分别与actor网络和critic网络相同。在进行Q值估计时采用目标函数值,即:

$$y_t=r+\gamma\min_{i=1,2}\left(Q_i'\left(s_{t+1},\mu'\left(s_{t+1}\middle|\theta_{\mu'}\right)\middle|\theta_{q_i'}+\varepsilon\right)\right) \tag{37}$$

式中:$\mu'\left(s_{t+1}\middle|\theta_{\mu'}\right)$为actor目标网络;$Q_i'\left(s,a\middle|\theta_{q_i'}\right)$为critic目标网络。

3)目标网络采用软更新(soft update)的方式使得参数缓慢变化,提高算法稳定性。软更新方式为:

$$\begin{cases}\theta_{\mu'}\leftarrow\tau\theta_\mu+(1-\tau)\theta_{\mu'}\\ \theta_{q_i'}\leftarrow\tau\theta_{q_i}+(1-\tau)\theta_{q_i'}\end{cases} \tag{38}$$

式中:τ为软更新系数,$0<\tau<1$;$i=1,2$。

4)TD3算法在critic网络进行一定次数的更新后再更新actor网络和目标网络的参数,延迟过高估计误差的传播,有利于将网络中的错误最小化。

TD3算法流程如附录A表A1所示,基于TD3算法的光储充电站储能系统优化流程如附录A图A1所示。

4 算例分析

4.1 算例设置

电动汽车充电需求数据来自美国La Canada国家研究实验室公开数据集[23];光伏数据来自Yulara光伏系统实际出力数据集[24];与电网交易的电价数据来自美国PJM市场实时电价数据集[25]。选取各数据集2018年4月至2020年4月共2 a的数据,将上述数据按每天96个时段进行处理得到本文算例的数据集,将其中1 a的数据作为训练集,其余数据作为测试集。

光储充电站系统的主要参数见附录B表B1,储能运行效率模型见附录B表B2[14],容量衰减模型参数见附录B表B3[16]。本文设定当电池容量衰减为额定容量的80%时,电池将无法正常使用[12]。本文假设储能能够在控温设备的作用下保持内部温度与参考温度相同,因此不考虑温度对储能容量衰减的影响。

本文采用TD3算法求解光储充电站储能优化运行问题。TD3算法参数设置见附录B表B4。actor网络的输入为光伏实际发电功率、电动汽车充电功率需求、电价和储能SOC,输出为动作,即储能充放电功率;critic网络输入为状态和动作,输出为该动作和状态下对应的Q值。actor网络和critic网络的隐层神经元为60×60×60,隐层激活函数选择leaky ReLU,TD3算法的神经网络结构图见附录B图B1。仿真测试平台的处理器为Intel Core i7-10700@2.90 GHz,采用16 GB RAM,DRL采用Python3.7和Tensorflow2.0架构进行编程和训练。

4.2 训练过程

模型训练回报曲线如图4所示。在前2 000个训练周期,经验回放池尚未存满,模型随机选取动作,对环境进行充分探索,不进行学习。在2 000个训练周期后,模型开始学习,逐步寻找最优策略,周

期回报逐渐增加，在约5 000个训练周期后，周期回报开始稳定，周期回报值收敛于15附近。储能在每个训练周期的惩罚值如图5所示。惩罚值在训练过程中逐渐减小，收敛于10左右，由于储能在每个周期都会有充放电动作，产生储能成本，所以惩罚不会减少到0，这说明TD3算法不仅能够收敛，还能够正确选择储能动作使得周期内的惩罚值最小。实际的回报曲线和惩罚曲线具有较大波动，有两方面原因：一是在训练过程中，为避免模型陷入局部最优，在选取动作时加入了随机噪声；二是电动汽车的充电需求、光伏出力都具有一定的不确定性，导致每个训练周期的环境状态可能具有较大差异，因此每个训练周期的回报值存在一定的差异性。

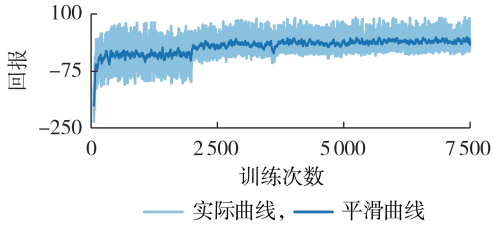


图4　训练过程回报曲线

Fig.4　Reward curves of training process

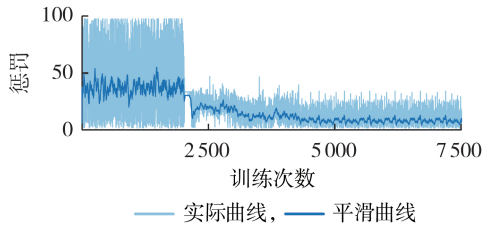


图5　训练过程惩罚曲线

Fig.5　Penalty curves of training process

考虑储能衰减模型和未考虑储能衰减模型的训练过程如图6所示。由图可见，未考虑储能衰减时，模型大约在3 500个训练周期后收敛，比考虑储能衰减时收敛更快，说明图3所示的储能衰减的计算步骤使得强化学习模型更难收敛。这是因为在训练过程中，强化学习智能体需要通过该计算步骤的结果α_k学习储能充放电动作所带来的容量衰减，进而更好地选择充放电策略，因此增加了强化学习的训练难度。由于未考虑储能的衰减成本，因此图6中未考虑储能衰减模型时的回报更高。本文算例设置的最大训练次数为7 500，在此情况下，2种方法都能收敛。因此，虽然储能衰减容量的计算步骤会增大强化学习的训练难度，但是并不会增加本文算例所用的时间成本。

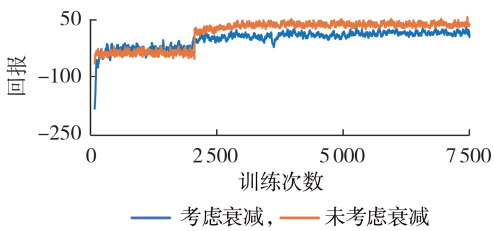


图6　训练过程对比

Fig.6　Training process comparison

4.3　算法测试

本文利用训练好的强化学习模型测试光储充电站储能系统的全寿命周期运行情况。为分析本文所提模型和算法的有效性，共选取了4种方法进行对比。其中，方法1为本文所采用的TD3算法，在训练时采用96点实际数据，在测试时只获取当前时段数据；方法2假设已知96点实际数据，求解储能充放电策略，因此方法2为理想情况；方法3采用“实际数据+15%的正态分布偏差数据”求解当前时段的储能充放电策略，该方法考虑了光伏发电、电动汽车充电以及电价的不确定性，以验证本文所提方法的鲁棒性；方法4未考虑储能损耗，同样假设已知96点实际数据，求解储能充放电策略。4种方法计算方式如表1所示，其中方法2—4采用Yalmip和CPLEX求解储能充放电策略，且将储能的运行效率设为常数0.98。由于本文在计算储能容量衰减成本时对储能的所有成本进行了折算，因此，本文将年平均收益作为判断优化效果优劣的标准。

表1　4种方法对比

Table 1　Comparison of four methods

方法	数据类型	信息知晓情况	是否考虑储能损耗
1	实际数据	当前时段数据	是
2	实际数据	优化周期所有数据	是
3	实际数据+15%的正态分布偏差数据	优化周期所有数据	是
4	实际数据	优化周期所有数据	否

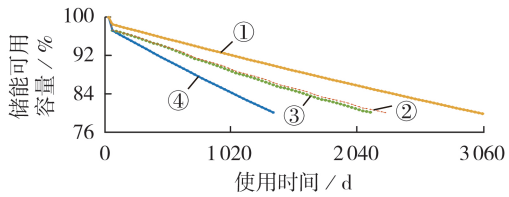
不同方法得到的光储充电站的收益如表2所示，4种方法下储能容量衰减曲线如图7所示。其中，方法1的储能运行时间最长，为3 060 d，储能全寿命周期的光伏充电站的收益最高，年平均收益也较高。方法2在优化时已知所有时段实际数据，为理想情况，因此方法2的优化效果最好，获得最高的年平均收益，但方法2需要获得优化周期内的实际数据，在实际运行中难以实现。方法3采用“实际数据+15%的正态分布偏差数据”，由于采用数据与实际数据存在一定偏差，从而导致方法3所得结果的全寿命周期收益和年平均收益都比方法2低。而方法1在只知道当前时段的数据的情况下，仍能比方法3获得更高的年平均收益，说明方法1所采用的TD3算法能有效解决不确定性环境下的储能充放电决策问题。方法4在不考虑储能动作损耗的情况下求解储能充放电功率和SOC，再通过图3中的储能

容量计算方法得到储能寿命，所获得的储能寿命最短。由此可知，在决定储能充放电策略时，如果不考虑储能容量损耗，会导致储能频繁充放电，使得储能寿命衰减加速，降低系统长期收益。

表2 光储充电站收益

Table 2 Revenues of photovoltaic-storage charging station

方法	储能寿命/d	全寿命周期收益/元	年平均收益/(元·a^{-1})
1	3060	1753316.87	209137.47
2	2340	1572462.01	235082.82
3	1950	1021376.45	191180.72
4	1350	635277.00	171760.08



① 方法1，② 方法2，③ 方法3，④ 方法4

图7 储能容量衰减曲线

Fig.7 Curves of energy storage capacity degradation

测试结果中某一天的储能SOC如图8所示，储能的充放电动作见附录C图C1。方法1中，储能可以在电价较高时放电，在电价较低时充电，通过电价差使得光储充电站获取收益，这说明TD3算法能够学习到使得收益更高的储能充放电策略。相比于未考虑储能损耗的方法4，考虑损耗的方法1、2减少了充放电的次数，降低了充放电的功率。相比于方法2和4，方法1所得到的储能充放电次数更少，这是由于为了延长储能寿命，在强化学习的奖惩函数式(27)中将储能成本的权重系数设置得较大，使得在决策时，方法1偏向于减少部分带来即时收益的动作，以代价更小的短期利益换取长期利益的增加。

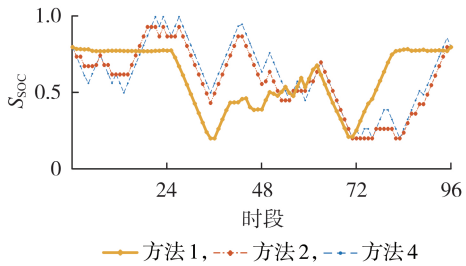


图8 储能SOC

Fig.8 SOC of energy storage

4.4 灵敏度分析

为进一步说明奖惩函数式(27)中权重系数对强化学习算法结果的影响，本文以储能容量衰减惩罚对应的权重系数σ_2为例进行灵敏度分析。不同σ_2取值下的储能寿命及光储充电站的收益情况如表3所示。

表3 权重系数σ_2灵敏度分析

Table 3 Sensitivity analysis of weight coefficient σ_2

σ_2	储能寿命/d	全寿命周期收益/元	年平均收益/(元·a^{-1})
1	1470	628930.40	156162.99
2.5	2640	1423569.44	196819.26
5	3060	1753316.87	209137.47
10	3510	1855964.82	192999.19
20	6210	2659745.21	156329.63

当$\sigma_2<5$时，储能容量衰减惩罚的权重系数小于储能末时段惩罚的权重系数，智能体在选择动作时会更加倾向于避免储能末时段惩罚的动作，通过相对频繁且幅度较小的充放电动作，使得SOC可以一直在初始时段的SOC S^1_{SOC}附近上下波动，同时获取一定的收益。随着σ_2的增大，储能的动作和获利方式更加合理，寿命和收益都增加。当$\sigma_2>5$时，智能体在决策时对储能动作所带来的容量衰减更加重视。随着σ_2的增大，智能体越来越倾向于减少储能的动作次数，以延缓容量衰减，从而获取较长的运行寿命和较高的全寿命周期收益。

本文在算例仿真中选用的包括σ_2在内的参数，是经过多次仿真实验后选择的一组优化结果较优的实验参数，可能并不是最优的，这也体现了神经网络的调参困难的缺点。此外，不同方法的计算时间如表4所示。方法1虽然需要较长的训练时间，但在模型训练好后的测试中运算速度快，每个时段的优化时间仅需0.0003 s左右且不受问题复杂度的影响，满足实时调度的要求。方法2—4不需要提前训练且都将储能的运行效率视为常数，但耗时显著长于方法1，这说明DRL算法在复杂问题中仍具有较高的计算效率。其中，方法2、3测试时间和平均优化时间最长。方法4未考虑储能衰减模型，耗费时间相较于方法2、3更少。

表4 4种方法的计算时间对比

Table 4 Comparison of calculating time among four methods

方法	训练时间/s	测试时间/s	各时段平均优化时间/s
1	9375	88.13	0.0003
2	0	32774.98	0.1459
3	0	28398.24	0.1517
4	0	1451.52	0.0112

5 结论

本文建立了考虑储能运行效率和容量衰减的储能优化运行问题，并采用双延迟深度确定性策略梯度算法进行求解，有效考虑了光伏发电、电动汽车充电负荷以及电价的不确定性。与现有方法相比，所

提出的模型和方法具有以下优势：

1）DRL无需获取未来时段实际数据就能得到较优结果，降低了预测误差对优化结果的影响；

2）DRL避免直接对不确定性建模，降低了光伏出力、电动汽车充电需求和电价等不确定性因素对优化结果的影响；

3）本文所建模问题考虑了储能容量衰减，使得光储充电站在储能全寿命周期的收益增加。

另一方面，由于在求解DRL问题时引入了神经网络，因此DRL具有调参困难、可解释性不高的缺点。另外，本文主要研究光储充电站中储能系统充放电策略，因此将所有电动汽车充电需求作为整体考虑，未研究每辆电动汽车的充电策略。考虑电动汽车的时空分布，优化电动汽车充电策略将是下一步的研究工作。

附录见本刊网络版（http://www.epae.cn）。

参考文献：

[1] 南琦琦，穆云飞，董晓红，等. 电动汽车快速充电网综合评估指标体系与方法[J]. 电力系统自动化，2020，44(1)：83-91.
NAN Qiqi，MU Yunfei，DONG Xiaohong，et al. Comprehensive evaluation index system and method for fast charging network of electric vehicles[J]. Automation of Electric Power Systems，2020，44(1)：83-91.

[2] 李景丽，时永凯，张琳娟，等. 考虑电动汽车有序充电的光储充电站储能容量优化策略[J]. 电力系统保护与控制，2021，49(7)：94-102.
LI Jingli，SHI Yongkai，ZHANG Linjuan，et al. Optimization strategy for the energy storage capacity of a charging station with photovoltaic and energy storage considering orderly charging of electric vehicles[J]. Power System Protection and Control，2021，49(7)：94-102.

[3] 王守相，张善涛，王凯，等. 计及分时电价下用户需求响应的分布式储能多目标优化运行[J]. 电力自动化设备，2020，40(1)：125-132.
WANG Shouxiang，ZHANG Shantao，WANG Kai，et al. Multi-objective optimal operation of distributed energy storage considering user demand response under time-of-use price[J]. Electric Power Automation Equipment，2020，40(1)：125-132.

[4] 胡代豪，郭力，刘一欣，等. 计及光储快充一体站的配电网随机-鲁棒混合优化调度[J]. 电网技术，2021，45(2)：507-519.
HU Daihao，GUO Li，LIU Yixin，et al. Stochastic / robust hybrid optimal dispatching of distribution networks considering fast charging stations with photovoltaic and energy storage[J]. Power System Technology，2021，45(2)：507-519.

[5] LIAO Y T，LU C N. Dispatch of EV charging station energy resources for sustainable mobility[J]. IEEE Transactions on Transportation Electrification，2015，1(1)：86-93.

[6] 路欣怡，刘念，陈征，等. 电动汽车光伏充电站的多目标优化调度方法[J]. 电工技术学报，2014，29(8)：46-56.
LU Xinyi，LIU Nian，CHEN Zheng，et al. Multi-objective optimal scheduling for PV-assisted charging station of electric vehicles[J]. Transactions of China Electrotechnical Society，2014，29(8)：46-56.

[7] 鲁卓欣，徐潇源，严正，等. 不确定性环境下数据驱动的电力系统优化调度方法综述[J]. 电力系统自动化，2020，44(21)：172-183.
LU Zhuoxin，XU Xiaoyuan，YAN Zheng，et al. Overview on data-driven optimal scheduling methods of power system in uncertain environment[J]. Automation of Electric Power Systems，2020，44(21)：172-183.

[8] 赵峰，李建霞，高锋阳. 考虑不确定性的高速公路光储充电站选址定容[J]. 电力自动化设备，2021，41(8)：111-117.
ZHAO Feng，LI Jianxia，GAO Fengyang. Siting and sizing of photovoltaic-storage charging stations on highway considering uncertainties[J]. Electric Power Automation Equipment，2021，41(8)：111-117.

[9] 雷金勇，郭祚刚，陈聪，等. 考虑不确定性及电/热储能的综合能源系统两阶段规划-运行联合优化方法[J]. 电力自动化设备，2019，39(8)：169-175.
LEI Jinyong，GUO Zuogang，CHEN Cong，et al. Two-stage planning-operation co-optimization of IES considering uncertainty and electrica / thermal energy storage[J]. Electric Power Automation Equipment，2019，39(8)：169-175.

[10] 季宇，熊雄，寇凌峰，等. 基于经济运行模型的储能系统投资效益分析[J]. 电力系统保护与控制，2020，48(4)：143-150.
JI Yu，XIONG Xiong，KOU Lingfeng，et al. Analysis of energy storage system investment benefit based on economic operation model[J]. Power System Protection and Control，2020，48(4)：143-150.

[11] 赵乙潼，王慧芳，何奔腾，等. 面向用户侧的电池储能配置与运行优化策略[J]. 电力系统自动化，2020，44(6)：121-128.
ZHAO Yitong，WANG Huifang，HE Benteng，et al. Optimization strategy of configuration and operation for user-side battery energy storage[J]. Automation of Electric Power Systems，2020，44(6)：121-128.

[12] 贺鸿杰，张宁，杜尔顺，等. 电网侧大规模电化学储能运行效率及寿命衰减建模方法综述[J]. 电力系统自动化，2020，44(12)：193-207.
HE Hongjie，ZHANG Ning，DU Ershun，et al. Review on modeling method for operation efficiency and lifespan decay of large-scale electrochemical energy storage on power grid side[J]. Automation of Electric Power Systems，2020，44(12)：193-207.

[13] KIM T，QIAO W. A hybrid battery model capable of capturing dynamic circuit characteristics and nonlinear capacity effects[J]. IEEE Transactions on Energy Conversion，2011，26(4)：1172-1180.

[14] MORSTYN T，HREDZAK B，AGUILERA R P，et al. Model predictive control for distributed microgrid battery energy storage systems[J]. IEEE Transactions on Control Systems Technology，2018，26(3)：1107-1114.

[15] 赵伟，袁锡莲，周宜行，等. 考虑运行寿命内经济性最优的梯次电池储能系统容量配置方法[J]. 电力系统保护与控制，2021，49(12)：16-24.
ZHAO Wei，YUAN Xilian，ZHOU Yixing，et al. Capacity configuration method of a second-use battery energy storage system considering economic optimization within service life[J]. Power System Protection and Control，2021，49(12)：16-24.

[16] XU Bolun，OUDALOV A，ULBIG A，et al. Modeling of lithium-ion battery degradation for cell life assessment[J]. IEEE Transactions on Smart Grid，2018，9(2)：1131-1140.

[17] CAO Jun，HARROLD D，FAN Zhong，et al. Deep reinforcement learning-based energy storage arbitrage with accurate lithium-ion battery degradation model[J]. IEEE Transactions on Smart Grid，2020，11(5)：4513-4521.

[18] 李航，李国杰，汪可友. 基于深度强化学习的电动汽车实时调度策略[J]. 电力系统自动化，2020，44(22)：161-167.
LI Hang，LI Guojie，WANG Keyou. Real-time dispatch strategy for electric vehicles based on deep reinforcement learning[J].

Automation of Electric Power Systems,2020,44(22):161-167.
[19] CHEN M,RINCON-MORA G A. Accurate electrical battery model capable of predicting runtime and *I-V* performance[J]. IEEE Transactions on Energy Conversion,2006,21(2):504-511.
[20] KE Xinda, LU Ning, JIN Chunlian. Control and size energy storage systems for managing energy imbalance of variable generation resources[J]. IEEE Transactions on Sustainable Energy,2015,6(1):70-78.
[21] LILLICRAP T P, HUNT J J, PRITZEL A, et al. Continuous control with deep reinforcement learning[EB/OL]. (2019-07-05)[2021-05-20]. https://arxiv.org/abs/1509.02971 arXiv preprint arXiv:1509.02971,2015.
[22] FUJIMOTO S,HOOF H,MEGER D. Addressing function approximation error in actor-critic methods[C]//Proceedings of the 35th International Conference on Machine Learning. Stockholm,Sweden:PMLR,2018:1587-1596.
[23] LEE Z J,LI Tongxi,LOW S H. ACN-data:analysis and applications of an open EV charging dataset[C]//Proceedings of the Tenth ACM International Conference on Future Energy Systems. Phoenix,AZ,USA:ACM,2019:139-149.
[24] DKA Solar Centre. Download data[EB/OL]. [2021-06-02]. http://dkasolarcentre.com.au/download.
[25] PJM. Download data[EB/OL]. [2021-06-02]. http://dataminer2.pjm.com.

作者简介:

陈亭轩

陈亭轩(1998—),女,重庆人,硕士研究生,主要研究方向为电力系统优化运行(**E-mail**:Chentingxuan@sjtu.edu.cn);

徐潇源(1989—),男,江苏泰兴人,助理教授,博士,通信作者,研究方向为电力系统不确定性分析、电力系统优化运行(**E-mail**:xuxiaoyuan@sjtu.edu.cn);

严 正(1964—),男,江西上犹人,教授,博士研究生导师,主要研究方向为电力系统优化运行、电力系统稳定分析及智能电网(**E-mail**:yanz@sjtu.edu.cn)。

(编辑 任思思)

Optimal operation based on deep reinforcement learning for energy storage system in photovoltaic-storage charging station

CHEN Tingxuan,XU Xiaoyuan,YAN Zheng,ZHU Yanming
(Key Laboratory of Control of Power Transmission and Conversion,Ministry of Education,
Shanghai Jiao Tong University,Shanghai 200240,China)

Abstract: Optimizing the energy storage charging and discharging strategy of photovoltaic-storage charging stations is conducive to improving the economics of system operation,but the existing model-driven stochastic optimization methods cannot fully consider the accurate energy storage system operating characteristics and the uncertainty of photovoltaic power generation and electric vehicle charging load. In this regard,an optimal operation method based on deep reinforcement learning for the entire life cycle of energy storage system in photovoltaic-storage charging station is proposed. Firstly,the refined model of energy storage operation efficiency and capacity degradation are modeled. Then considering the uncertainty of electric vehicle charging demand,photovoltaic output and electricity price,under the condition of meeting electric vehicle charging demand and photovoltaic consumption,an optimal operation method for energy storage based on reinforcement learning is established,which takes maximizing the total revenue of photovoltaic-storage charging station as its target. Considering the action continuity of the energy storage charging and discharging decision-making,the twin delayed deep deterministic strategy gradient algorithm is used to solve the problem. The historical data is used to train the model,and the energy storage charging and discharging strategy can be optimized in real time according to the current state. Finally,the proposed method and model are tested and compared with the traditional model-driven methods,the results verify the effectiveness of the proposed method and model.

Key words: energy storage; photovoltaic-storage charging station; uncertainty; deep reinforcement learning; optimization

附录 A:TD3 算法流程

表 A1 TD3 算法流程

Table A1 Flowchart of TD3 algorithm

TD3 算法
初始化网络模型参数，初始化强化学习环境
设置最大训练次数 ep_max，周期步数 T_0，样本数 m，延迟步数 d
For ep=0:ep_max do
For i=0:T_0 do
$j \leftarrow j+1$
获取环境状态 s_t，按照式（33）产生动作 a_t
执行 a_t 并返回 r_t 和 s_{t+1}
将样本 (s_t, a_t, r_t, s_{t+1})存入经验回放池 D
从 D 中随机采样 m 个样本，通过最小化式（35）更新 critic 网络参数
If j 为 d 的整倍数 then
通过式（34）更新 actor 网络参数
通过式（38）更新目标网络参数
End If
End For
End For

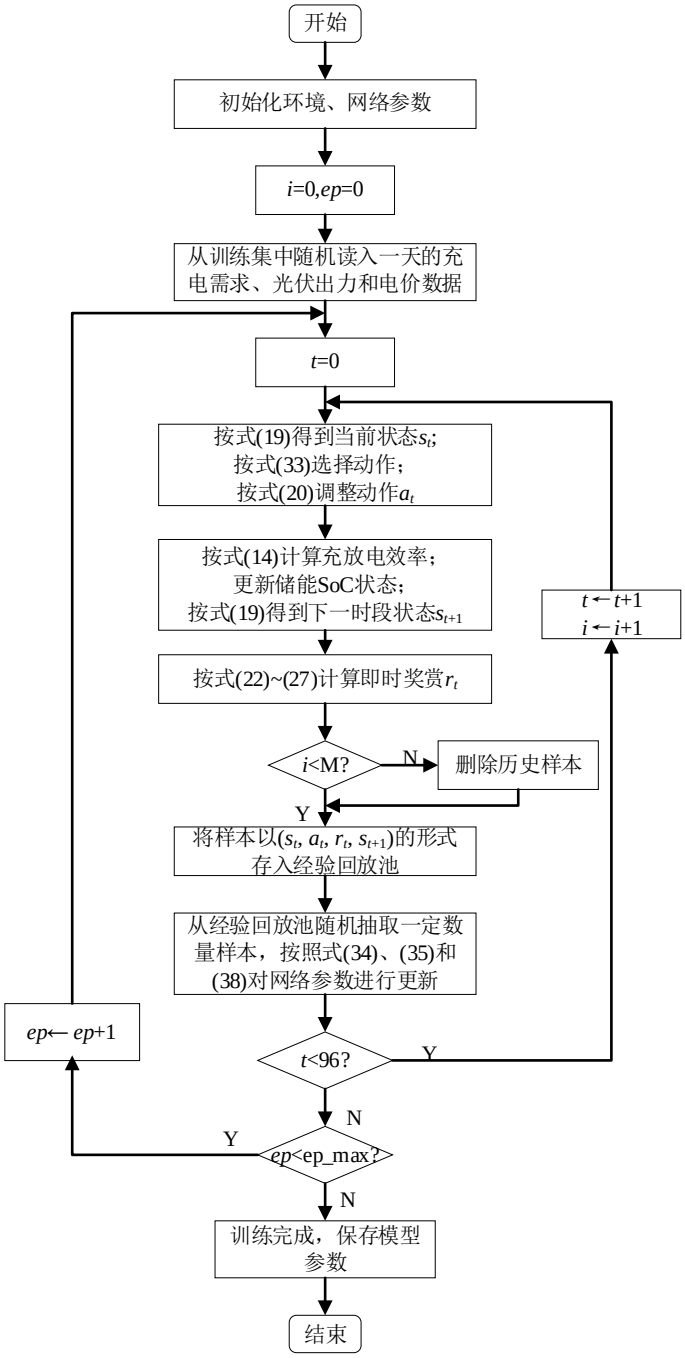


图 A1 基于 TD3 算法的储能优化运行问题求解流程

Fig.A1 Flowchart of solving energy storage system operation problem

附录 B:算例相关参数

表 B1 光储充电站系统参数

Table B1 Parameters of solar-powered charging station system

参数	数值
$P_{es,max}$/kW	100
E_{ini}/kW	400
S_{SOC}^{min}	0.20
S_{SOC}^{max}	1.00
Δt/h	0.25
T_0	96

表 B2 储能运行效率模型相关参数

Table B2 Parameters of energy storage efficiency model

参数	数值	参数	数值
e_0	1.00	e_1	4.00×10^{-3}
e_2	-3.11×10^{-3}	e_3	-4.77×10^{-3}
e_4	3.06×10^{-3}	e_5	9.66×10^{-8}
h_0	1.00	h_1	-4.60×10^{-3}
h_2	4.13×10^{-3}	h_3	5.00×10^{-7}
h_4	4.23×10^{-13}	h_5	-1.36×10^{-7}

表 B3 储能容量衰减模型相关参数

Table B3 Parameters of energy storage capacity degradation model

参数	数值	参数	数值
α_{sei}	0.0575	β_{sei}	121
$k_{\delta 1}$	140000	$k_{\delta 2}$	-0.501
$k_{\delta 3}$	-123000	k_{SOC}	1.04
$S_{SOC,ref}$	0.5	k_T	0.0693
T_{ref}/℃	25	k_t/s	4.14×10^{-10}

表 B4 TD3 算法参数

Table B4 Parameters of TD3 algorithm

参数	数值	参数	数值
状态空间维度	4	延迟步数 d	100
动作空间维度	1	最大训练次数	7500
策略网络学习率	0.005	权重系数 σ_1	1
值网络学习率	0.0005	权重系数 σ_2	5
采样容量 m	128	权重系数 σ_3	5
经验回放池容量	19200	权重系数 σ_4	5

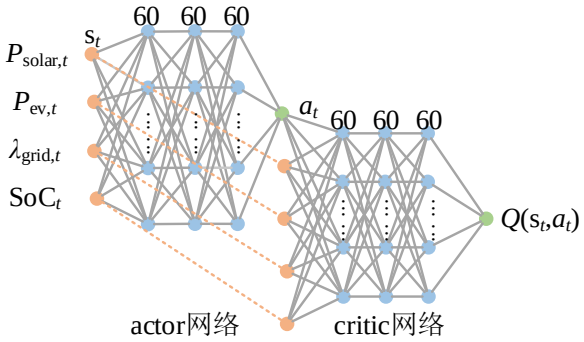


图 B1 神经网络结构图

Fig.B1 Neural network architecture

附录 C:部分算例结果

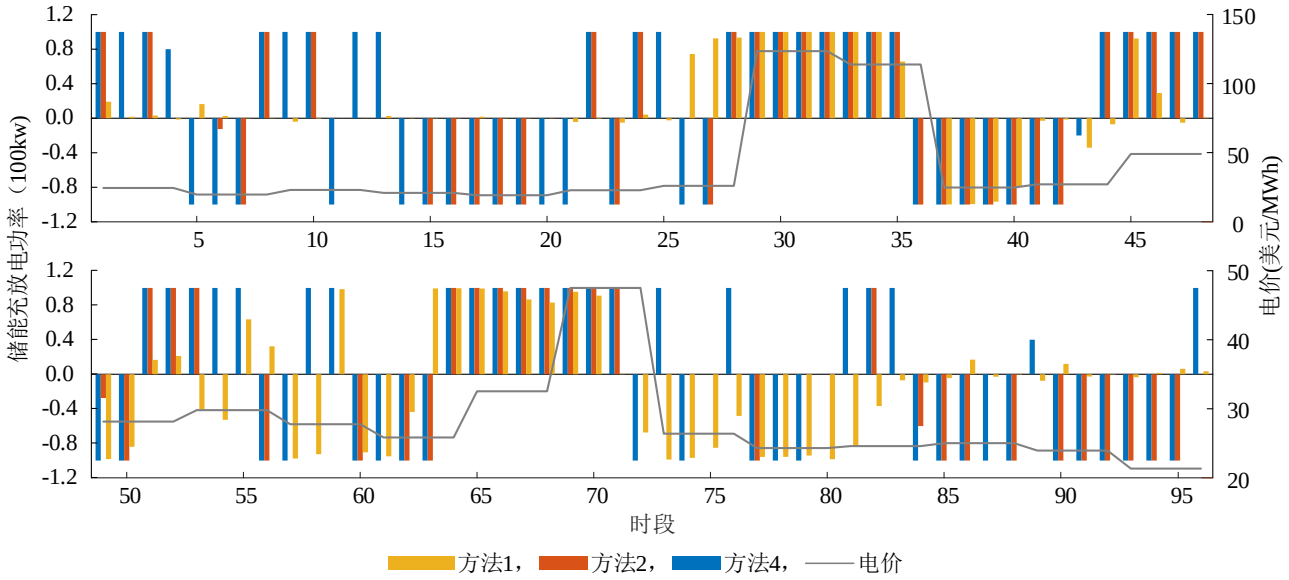


图 C1 储能充放电功率

Fig.C1 Charging/discharging power of energy storage